DOT/FAA/AM-96/22

Office of Aviation Medicine
Washington, D.C. 20591

A Further Validation of the Practical Color Vision Test for En Route Air Traffic Control Applicants

H.W. Mertens
N.J. Milburn
W.E. Collins

Civil Aeromedical Institute
Federal Aviation Administration
Oklahoma City, Oklahoma 73125

August 1996

Final Report

This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

U.S. Department of Transportation

Federal Aviation Administration

NOTICE

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/AM-96/22		

4. Title and Subtitle	5. Report Date
A Further Validation of the Practical Color Vision Test for En Route Air Traffic Control Applicants	August 1996
	6. Performing Organization Code
7. Author(s) H.W. Mertens, N.J. Milburn, and W.E. Collins	8. Performing Organization Report No.
9. Performing Organization Name and Address FAA Civil Aeromedical Institute P.O. Box 25082 Oklahoma City, OK 73125	10. Work Unit No. (TRAIS)
	11. Contract or Grant No.
12. Sponsoring Agency name and Address Office of Aviation Medicine Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591	13. Type of Report and Period Covered
	14. Sponsoring Agency Code

15. Supplemental Notes

16. Abstract

The Flight Progress Strips Test (FPST) is currently used for secondary color vision screening of applicants for air traffic control jobs at en route centers. The test provides a practical, job-specific color vision selection criterion involving use of color coding in the most important color task of en route radar contollers, i.e., discrimination of the non-redundant color coding in flight progress strips (FPSs). This experiment provides a further, independent validation of the FPST using a new criterion test. Prediction by the FPST of performance on the new and old criterion tests was compared. Subjects were classified as normal or deficient based on anomaloscope readings. The pass/fail cutoff score for all tests was "pass with no more than one error." All people with normal color vision passed. Over all, for participants with both normal and abnormal color vision, the correlations between error scores on the FPST and both criterion tests were greater than r=.93, and error scores tended to increase with degree of color vision deficiency. The validity of the FPST was Kappa=.86 for prediction of performance on the new criterion test, as compared to .91 for prediction of performance on the original criterion test. Part of that small decrease in validity may be because of application of the same pass/fail cutoff score to the new criterion test, which contains a larger number of items than the FPST. The predictive validity of the FPST was shown to be acceptably high in this further validation with a new, independent set of actual flight progress strips as the criterion test.

17. Key Words	18. Distribution Statement
Air Traffic Controllers Color vision standards Performance tests Color vision tests	Document is available to the public through the National Technical Information Service Springfield, Virginia 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	12	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

A FURTHER VALIDATION OF THE PRACTICAL COLOR VISION TEST FOR EN ROUTE AIR TRAFFIC CONTROL APPLICANTS

INTRODUCTION

Practical, "job-sample" color vision tests were developed by the Federal Aviation Administration's (FAA's) Office of Aviation Medicine to optimize the validity and fairness of color vision testing for air traffic control specialist (ATCS) selection (Mertens, Milburn, & Collins, 1995). The two practical color vision tests, one for the en route center work option and another for work at terminal facilities, evaluate the ability to perform the most important, safety-critical tasks involving color coding in each work situation. They have been implemented for secondary color vision screening of ATCS applicants who fail clinical color vision screening during their pre-employment medical examination by an Aviation Medical Examiner. The present research was conducted to obtain additional evidence of the validity of the practical color vision test for the en route center work option.

The most demanding color vision task at en route centers involves accurately distinguishing between red and black computer printing and handwriting on Flight Progress Strips (FPSs). The test for this task is named the Flight Progress Strips Test (FPST). There is at least one FPS for each aircraft that is controlled by the en route ATCS, and the FPSs are frequently referred to during an ATCS's time "on the boards" (Vortac, Manning, & Rotter, 1992). The FPS is a paper strip, 33.3 mm by 203 mm, and contains a variety of information that is either computer printed by the air traffic control (ATC) system's computers, or handwritten by the ATCSs. The most critical color coding on the FPS involves the use of the colors black and red to differentiate between (1) assigned versus non-assigned (planned) information, (2) coordination versus incorrect reporting altitude, (3) north versus south flights, and (4) east versus west flights. The red/black color coding is typically non-redundant, especially in the critical case of distinguishing assigned from non-assigned information concerning altitude, route, departure, approach, fixes, clearances, and so forth (FAA, 1989). Misidentification of that critical information about an aircraft's status can result in danger to that aircraft or place it in conflict with another aircraft.

The FPST requires the identification of red and black colors in reproductions of actual FPSs made by a photographic color separation process and a lithographic reproduction process that are also used for color-critical fine art printing. The FPS copies were printed on the same paper used by the FAA for FPSs. When the FPST is given, the test book is placed on a special stand containing a light source developed to produce a standard test illumination, with the color and illumination level typical of actual ATCS work stations.

The FPST was originally validated by directly comparing FPST color responses with similar responses to a criterion test which consisted of the actual strips from which the copies in the test were made. The original criterion test is referred to as Criterion Flight Progress Strips Test-1 (CFPST-1). The criterion for passing both FPST and CFPST-1 was "no more than one error." That criterion was based on previous research with similar materials in which normal trichromats made no more than one error (Mertens & Milburn, 1992a). The predictive validity of the FPST, with CFPST-1 as the criterion, was high (Kappa=.91). The reliability of the FPST was also high (alpha=.93). The high job-relevance, validity, and reliability of the FPST increases the effectiveness and fairness of ATCS color vision testing.

The present research provides additional evidence of FPST validity by evaluating it's ability to predict color responses in a new criterion test. This independent verification of validity also demonstrates that FPST-based predictions of ability to discriminate and identify colors in FPS materials can be generalized beyond the original criterion test to a larger body of FPSs created daily in operations of the FAA's en route centers.

METHODS

Participants

The tests were evaluated using the same, 191 people who participated in the original validation study (Mertens, Milburn, & Collins, 1995). Normals included 63 men and 43 women. Color deficients included 84 men and 1 woman. (The incidence of inherited CVD is less than ½ of 1% of all women compared to 8 to 10% of all men.) The ages of participants with normal color vision ranged from 18 to 58 yr with a mean age of 31.2 yr and a standard deviation of 11.1 yr. The ages of participants with color vision deficiency (CVD) ranged from 18 to 61 yr with a mean age of 35.0 yr and a standard deviation of 12.5 yr.

All participants had at least 20/30 visual acuity in both near and distant vision, as measured by the Bausch and Lomb Orthorater vision tester. Participants were recruited through advertisements in newsletters at local colleges and universities, and through local newspapers of the Oklahoma City (OK) metropolitan area. All were paid an hourly wage.

FPST and CFPST-1

The FPST included a copy of each of 30 FPSs selected from many thousands of strips obtained from six en route centers; half of one day's output of strips was obtained from each center. The 30 original strips were copied to make the FPST, and then the original strips were used to make CFPST-1. The 30 FPSs were selected to be representative of the type fonts of the two types of computer printers that are used, the types of pens and pencils used for handwriting on strips, and representative of the range of darkness/lightness normally found in both computer printing and handwriting in the actual work environment.

Each group of computer-printed characters in the same line and column of the FPS, that did not contain a blank space and were the same color, were considered an item on the test section pertaining to computer printing. The six strips of the computer printing section had 100 items. Each handwritten altitude in the altitude column of a strip was an item in the section testing color vision for handwriting on strips; there were 24 items in this section of the test. The FPS copies and instructions for the test were assembled in book form to make the FPST. Three of the 30 strips were used for demonstration and 27 served as test items. Respondents identified the colors red and black in computer printing on 6 strips containing 100 items and the color of handwriting on 21 strips containing 24 items. There was no time limit for responding and performance was assessed in terms of total number of errors and a pass-fail score. An error consisted of misidentifying the color of a red or black item on a strip. The criterion for passing was "no more than one error." The CFPST-1 and FPST were essentially identical with regard to form of the test book, administration procedure, illumination, and scoring. The individual strips in each of the two parts of the FPST and CFPST-1, the computer printing and handwriting sections, were presented in different random orders in the two tests. Because the same strips and their copies were involved in the CFPST-1 and the FPST, the total number of items was identical on those two tests.

CFPST-2

The CFPST-2 included 30 new FPSs, 3 for demonstration, 6 for color discrimination in computer printing, and 21 for color discrimination in handwriting. Again, they were selected from many thousands of new strips obtained from the same six en route centers. The criteria for selection of representative strips were identical to the criteria, given above, for preparation of the CFPST-1. The six strips of the computer printing section of the CFPST-2 had 110 items. The 21 strips in the handwriting section contained a total of 28 items.

The same illuminant/test holder was used for all three tests. It was designed to hold the test books in a standard, convenient testing position, and to provide appropriate, standard illumination for the CFPST-1, CFPST-2, and FPST. The design is described more completely elsewhere (Mertens, Milburn, & Collins, 1995). A small cool white fluorescent lamp provided an illumination level of 59 lux at the center of the test page. That corresponded to the average workstation illumination as measured at two facilities, the Ft. Worth (TX) and Kansas City (KS) en route centers.

Diagnostic Color Vision Tests

The procedure used for classification of deficiencies utilized the Nagel Type I anomaloscope to classify individuals with red-green CVD. Other tests were used to detect and diagnose the rare blue-yellow deficiencies that the Nagel anomaloscope does not detect, but none of the latter were found. These diagnostic methods were identical to those described by Mertens and Milburn (1992a).

Procedure

The FPST, CFPST-1, and CFPST-2 were administered in the same test session by a trained laboratory technician. The order in which the three tests were given was counter balanced over all participants in the following manner. The CFPST-1 and FPST were given with the order reversed for half of the participants. The CFPST-2 preceded them in half the participants and followed them in half the participants. The three tests took approximately 40 minutes to administer. Several color vision tests were given at three other stations prior to or following FPS testing. Those other color vision tests were given for visual acuity screening, classification of CVD, and another color vision study. All testing of a particular subject was performed in a three-hour test session. Within each session, four test periods were separated by three 5-10 min breaks.

RESULTS AND DISCUSSION

Diagnostic Classification of Participants

The number of participants in each classification of type and degree of CVD are shown in Table 1. A complete description of anomaloscope classifications and criteria for each is available in Mertens and Milburn (1992a).

Representation of all categories of red-green CVD was obtained. The number of normal trichromats was 106 and the total number of deficients was 85. There were 35 deficients who were protans (pro) and 50 deficients who were deutans (deu).

Relation of Performance on Flight Progress Strips Tests to Color Vision Classification

Pass/Fail Performance. The CFPST-1, CFPST-2, and FPST tests permit comparison of responses between two sets of actual flight progress strips, and between the color copies in the FPST and two sets of actual strips. The criterion for passing each of the three tests was "no more than one error". That criterion was based on use of the similar FPS materials in a previous experiment (Mertens & Milburn, 1992a) in which no normal trichromat made more than one error. The relationship of passing and failing to type and degree of CVD is shown in Table 2.

TABLE 1

Anomaloscope Classifications of Subjects

Normal	Anomalous Trichromat				Dichromat	
	Simple		Extreme			
Trichromat	Pro	Deu	Pro	Deu	Pro	Deu
106	11	18	11	21	13	11

TABLE 2

Number of Subjects Passing and Failing on the CFPST-1, a CFPST-2, and FPST as a Function of Type and Degree of Color Vision Deficiency

		Normal Trichromat	Anomalous Trichromat Simple Pro	Anomalous Trichromat Simple Deu	Anomalous Trichromat Extreme Pro	Anomalous Trichromat Extreme Deu	Dichromat Pro	Dichromat Deu
CFPST-1	Pass	106	7	11	0	3	0	0
	Fail	0	4	7	11	18	13	11
CFPST-2	Pass	106	4	10	0	3	0	0
	Fail	0	7	8	11	18	13	11
FPST	Pass	106	7	9	0	5	0	0
	Fail	0	4	9	11	16	13	11

All individuals with normal color vision passed all three tests. Only one individual with normal color vision made one error; that was on CFPST-1. The probability of failing the flight progress strips tests increased with degree of CVD. Neither protans nor deutans who were dichromats were able to pass any test. Only a few (three) of the extreme anomalous trichromats were able to pass all three FPS tests, and all were deutans. Approximately half of both simple protanomalous and simple deuteranomalous trichromats, who have mild-to-moderate deficiencies, were able to pass all tests. Overall, the impact of CVD on color discrimination in FPSs seems comparable in both protans and deutans in terms of pass/fail scoring, but not in terms of total errors as discussed below. Simple anomalous trichromats had approximately a 50% chance of passing, extreme anomalous trichromats had approximately a 15% chance of passing, and no dichromats were able to pass.

Error Scores. The mean and standard deviation of total error scores and 95% confidence intervals for the mean are shown for the three FPS tests in Table 3 as a function of type and degree of CVD. The confidence intervals are presented for comparison of means for color deficient groups and normal trichromats. The low, near-zero variance among normal trichromats makes analysis of variance (ANOVA) inappropriate for comparison of the means of color deficient groups and normals (Box, 1954). Error scores for the two criterion tests and the FPST increase with degree of CVD, and the 95% confidence intervals for the means of all categories of deficients typically did not include the mean for normal trichromats; the exception was the small overlap of the confidence interval for the simple protanomalous category. That overlap was because of somewhat higher variance among the simple protanomalous on the FPST, rather than a change in mean errors.

While the mean error values tend to be slightly higher among the mild-to-moderate deutans (simple deuteranomalous) than for the milder protans (simple protanomalous), that difference does not seem of practical significance because of the great overlap of the distributions and confidence intervals for those groups. In contrast, the deutans with stronger deficiencies, (the extreme deuteranomalous and deuteranopes), make a substantially smaller number of errors, on the average, than the strong protans (extreme protanomalous and protanopes).

Analysis of Variance of Error Scores

Analyses of variance (ANOVA), with normal trichromats excluded, were performed with the MANOVA (Multiple Analysis of Variance) program

TABLE 3

Errors on FPST and CFPST as a Function of Type and Degree of Color Vision Deficiency

	Diagnosis	Mean	Standard Deviation	95% Confidence Limits Low	to	High
CFPST-1						
	Normal	0.01	0.10	0.01	to	0.03
	Protans					
	Simple	1.91	2.39	0.31	to	3.51
	Extreme	26.91	4.93	23.60	to	30.22
	Dichro.	29.54	7.58	24.96	to	34.12
	Deutans					
	Simple	3.50	5.83	0.60	to	6.40
	Extreme	7.67	4.20	5.76	to	9.58
	Dichro.	11.91	7.30	7.01	to	16.81
CFPST-2						
	Normal	0.00	0.00	0.00	to	0.00
	Protans					
	Simple	3.27	3.61	0.85	to	5.70
	Extreme	31.27	9.42	24.94	to	37.60
	Dichro.	34.53	15.87	24.95	to	44.13
	Deutans					
	Simple	3.44	5.17	0.87	to	6.02
	Extreme	6.76	5.75	4.14	to	9.38
	Dichro.	12.81	7.74	7.61	to	18.02
FPST						
	Normal	0.00	0.00	0.00	to	0.00
	Protans					
	Simple	2.73	4.94	-0.59	to	6.05
	Extreme	24.46	5.94	20.47	to	28.44
	Dichro.	27.92	9.12	22.42	to	33.43
	Deutans					
	Simple	2.94	5.35	0.28	to	5.61
	Extreme	5.48	5.07	3.17	to	7.78
	Dichro.	9.55	7.49	4.52	to	14.58

of the SPSS statistical system (SPSS, Inc., Chicago, IL). The error scores were evaluated with ANOVA as a function of type and degree of CVD as "between groups" factors and test (CFPST versus FPST) as a "within groups" factor. The assumption of homogeneity of variance was rejected by the Box-M test. Because ANOVA is robust for violation of this assumption when the number of participants in each group is equal, the method of compensating for heterogeneity of variance by randomly discarding cases to achieve equal cell numbers was used (Tabachnick & Fidell, 1989; Toothaker, 1986). While this same technique was used by Mertens and Milburn (Mertens, Milburn, & Collins, 1995) for comparison of FPST with CFPST-1 in the original validation, a new, independent random sample of participants was selected from each of the six groups with CVD for the present analysis.

Each of 6 groups, formed by the two types and three degrees of CVD, had 11 cases. The main effects, of both type and degree of CVD, were statistically significant. Protans made significantly more errors than deutans on the FPS tests [$F(1,54)=79.06$, $p<.001$], and errors increased significantly with degree of color vision deficiency [$F(2,54)=46.43$, $p<.001$], as discussed above. The interaction of type and degree was also significant [$F(2,54)=23.87$, $p<.001$]; there was greater difference between protans and deutans among extreme anomalous trichromats and dichromats, than among simple anomalous trichromats. There was also a significant difference for the main effect of test, i.e., the within groups factor [$F(2,108)=7.53$, $p=.001$]. No interaction involving the test factor was significant. The significant main effect of test reflects the fact that error scores tended to be slightly higher in the CFPST-1 test than in the FPST, and slightly higher in CFPST-2 than in CFPST-1. The latter difference may be related to the fact that more errors were possible in CFPST-2, because it had 14 more items. The dichromats and extreme anomalous groups averaged approximately two more errors on the FPST than the CFPST. However, mean errors for the simple anomalous trichromats were very similar in the two tests; the simple protanomalous actually had a mean of 0.82 more errors on the FPST, whereas the simple deuteranomalous had a mean of 0.55 more errors on the CFPST. As shown below, any small difference in difficulty between the two tests was not reflected in pass/fail performance. The number of participants who passed both tests was exactly the same (127), as shown in the marginal totals of Table 4.

Validity of the FPST

The validity of the FPST was evaluated by comparing pass/fail performance on FPST to pass/fail performance on the CFPST-1 and CFPST-2. Cohen's Kappa (k), an index of agreement, was used to assess the validity of the practical test, as recommended by the NRC-NAS Committee on Vision (1981). The index can be interpreted as the percentage agreement between test and criterion variable, with correction for chance. The relationship of passing and failing on the FPST, to passing and failing on both the CFPST-1 and CFPST-2 is shown in Table 4. The validity of the FPST, for prediction of pass/fail performance on the CFPST-2, was $k=.86$, compared to $k=.91$ with CFPST-1 as the criterion. While $k=.86$ is reasonably high, it should be considered that the CFPST-2 has more items.

There are two kinds of errors to consider regarding the FPST, *misses* (false negatives) and *false alarms* (false positives). The *miss rate* is the probability of passing the FPST, given that the criterion test was failed. The *false alarm rate* is the probability of failing the FPST, given that the criterion test was passed. If the criterion for passing CFPST-2 is changed from "no more than 1 error" to "no more than 2 errors", the validity increases to $k=.88$, and the miss rate is reduced from 12% to 8%, close to the 6% value obtained with CFPST-1, the original criterion test. Those error rates are considered acceptably low when used to predict the criterion of normal versus abnormal classification by an anomaloscope test, and are comparable to error rates of the better clinical color vision tests.

Reliability of the CFPST-2

Chronbach's alpha was computed as a conservative estimate of reliability, as used in previous study of the reliability of the FPST and CFPST-1. Chronbach's standardized item alpha, based on standardized item scores, was previously reported (Mertens, Milburn, &

TABLE 4

Comparison of Pass/Fail Performance on the CFPST-1, CFPST-2 and FPST

		CFPST-2	
		Pass	Fail
CFPST-1	Pass	122	5
	Fail	1	63
	Total	123	68
	Kappa=0.93		

		FPST	
		Pass	Fail
CFPST-1	Pass	123	4
	Fail	4	60
	Total	127	64
	Kappa=0.91		

		FPST	
		Pass	Fail
CFPST-2	Pass	119	4
	Fail	8	60
	Total	127	64
	Kappa=0.86		

Collins, 1995) as .93 for the FPST, and .95 for the CFPST-1. The Chronbach's standardized item alpha for CFPST-2 is .93. It can be concluded that both FPST and both criterion tests have high estimated reliability.

CONCLUSIONS

The present research is the fourth study, in a series of studies, on the ATCS color vision standard that has addressed the need for a color vision standard, the validity of FAA-accepted clinical color vision tests, and the development, evaluation, and validation of two practical color vision tests for ATCSs. The need for practical, "job-sample", color vision selection tests is evidenced by (1) the fact that there are some individuals with CVD who can perform the ATC color tasks as well as normals, and (2) some common clinical color vision tests have false alarm rates that are too high (Mertens & Milburn, 1992a, 1992b).

Optimally, color vision screening of ATCS applicants should provide both the ability to select all individuals who can perform safety-critical ATCS color tasks accurately, whether they have a CVD or not, and screen out all individuals who cannot perform those tasks accurately. A test with high predictive validity for performance in ATCS color tasks must have both a low miss (false negative) rate to ensure safety, and a low false alarm (false positive) rate to ensure fairness. The most direct way of accomplishing that objective is to create color vision screening tests that are "work sample" tests such that they accurately simulate the color coded materials of the task and the viewing conditions of the work environment. The stimulus characteristics that affect color perception, such as color, color contrast, brightness, brightness contrast, and stimulus size, must be reproduced accurately. Viewing conditions that can affect color perception, such as intensity and color of ambient illumination and viewing distance, must also be reproduced accurately. The practical tests should also simulate normal work situations, rather than more marginal observation conditions that sometimes occur, because those conditions may also degrade performance of individuals with normal color vision. Finally, the tests should involve very simple color identification and/or discrimination responses so that the tests assess only color vision ability. Such tests can be considered "practical tests" because they involve essentially the same materials and observation conditions that are involved in the actual work tasks (Jewell & Siegall, 1990).

The present study adds evidence of the high validity of the FPST. These results, together with findings of the initial validation study (Mertens, Milburn, & Collins, 1995), provide a data base supporting the FPST which shows that it is an effective and fair practical color vision test for screening en route ATCS applicants.

REFERENCES

Box, G.E.P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems. I. Effect of inequality of variance in the one-way classification. *Annals of Mathematical Statistics*, 25, 290-302.

Federal Aviation Administration. (1989). Air traffic control. (FAA Order 7110.65F). Washington, D.C.: Author.

Jewell, L.N., & Siegall, M. (1990). *Contemporary industrial and organizational psychology.* (2nd Ed.). St. Paul, MN: West Publishing Co.

Mertens, H.W., & Milburn, N.J. (1992). Performance of color-dependent tasks of air traffic control specialists as a function of type and degree of color vision deficiency. Report DOT/FAA/AM-92/28. Washington, D.C.: Federal Aviation Administration, Office of Aviation Medicine. (a)

Mertens, H.W., & Milburn, N.J. (1992). Validity of clinical color vision tests for air traffic control specialists. Report DOT/FAA/AM-92/29. Washington, D.C.: Federal Aviation Administration, Office of Aviation Medicine. (b)

Mertens, H.W., Milburn, N.J., & Collins, W.E. (1995). Practical color vision tests for air traffic control applicants: En route center and terminal facilities. Report DOT/AM-95/13. Washington, D.C.: Federal Aviation Administration, Office of Aviation Medicine.

National Research Council-National Academy of Sciences, Committee on Vision. (1981). *Procedures for testing color vision: Report of Working Group 41*. Washington, D.C.: National Academy Press.

Tabachnick, B.G., & Fidell, L.S. (1989). *Using multivariate statistics*, New York: Harper and Row. (p. 48-49)

Toothaker. L.E. (1986). *Introductory statistics for the behavioral sciences*. New York, McGraw-Hill. (p. 452)

Vortac, O.U., Manning, C.A., and Rotter, A.J. (1992) En route air traffic controller's use of flight progress strips: A graph-theoretic analysis. Report DOT/FAA/AM-92/31. Washington, D.C.: Federal Aviation Administration, Office of Aviation Medicine.

☆U.S. GOVERNMENT PRINTING OFFICE: 1996 -760-082/40007